We sincerely thank both reviewers for their insightful comments on our manuscript, which have greatly helped to clarify our findings. The main changes made to our manuscript include:

- The addition of three new figures illustrating the findings. New Figure 9 compares the PlankTOM10 and PlankTOM6 zonal mean biomass for the main phytoplankton and zooplankton PFTs and provides further background information on their differences; New Figure 10 shows the seasonal cycle in surface Chl in the North and South Pacific oceans and shows to which extent the modelled Chl seasonality in PlankTOM10 are closer to the observations than that those of PlankTOM6; New Figure 13 shows the biomass of phytoplankton and the three zooplankton in the North and South Pacific Ocean in the two models to illustrate the effect of the timing of zooplankton grazing on phytoplankton discussed in the text.

- We clarified a number of model results that were highlighted by the two reviewers, namely: our use of Chl and biomass, which is now more precise in the text, with a more extensive analysis of how grazing influences biomass; that our model does include it's own representation of mixed-layer dynamics, and thus it represents the vertical dilution effects that are present in the ocean; and more explanation and rationale for the choices of model parameters.

- We extended the discussion of the limitations of our modelling analysis, but also explained more clearly why we think the model is adequate to study the role of macrozooplankton grazing, in spite of the model shortcoming.

Please find below our reply (in blue) along with the new text (in red). The original comments (in black) are also copied for clarity.

V. Smetacek (Referee)

victor.smetacek@awi.de

The Southern Ocean (SO) plays a key role in the ocean carbon cycle hence efforts to model its past, current and possible future impact on atmospheric CO_2 levels are most worthy of discussion in a broad forum. This general rule applies particularly to this paper because the authors are prominent members of the ocean modelling community. Their message is that adding more zooplankton grazers, in this case , "large, slow-growing crustacean zooplankton", to a global ocean biogeochemistry model produces phytoplankton biomass levels in the SO closer to values obtained from satellite imagery. An earlier model without the additional zooplankton components gave unrealistically high summer phytoplankton biomass values in the SO. This is a logical conclusion: plant biomass will go down if the growth rate is kept constant but the grazing pressure increased. From this straight-forward balance equation the authors conclude that grazing rate rather than iron supply is responsible for the low chlorophyll concentrations in the Southern Ocean. Since this generalisation would bring the great HNLC debate of the 1990s - is it light, iron or grazing that controls productivity? - back to square one, it is necessary to review the arguments for the case made here in order to help clear up what is probably widespread confusion in the community regarding the extent of iron limitation in HNLC regions and the capabilities of biogeochemical models to deal with the issue.

Comment only – no reply needed.

A number of questions arose in my mind whilst reading the manuscript to which I could not find the answers in the manuscript and supplementary information. If the issues have been considered they should be mentioned prominently in the main text. If, on the other hand, one or more of these issues has not been considered, then the text will need to be rewritten in the appropriate places and the conclusions modified accordingly.

Comment only – no reply needed.

Question 1: Since phytoplankton biomass is equated with chlorophyll concentrations throughout the text, I would like to know whether the following factors that affect its variability have been considered and how: a) Latitudinal variation in mixed layer depth, b) increasing C/Chl ratios with declining iron supply.

a) Mixed layer depth (MLD) The ,"phytoplankton biomass" of interest to food web and carbon cycle studies is the integrated stock per area of water column (in mg or g chlorophyll or carbon m-2) and not just the concentration. The difference between concentration (obtained from discrete measurements) and stock (obtained from integrating discrete values for the mixed layer) is highlighted by the differences in chlorophyll yields between the OIF experiments SEEDS I in the Subarctic Pacific and EIFEX in the SO: >20 mg Chl m-3 and 3 mg Chl m-3 respectively. However, the mixed layer during SEEDS I was only 10 m deep but 100 m during EIFEX, so the standing stocks were 200 and 300 mg Chl m-2 respectively. As others have fallen into the MLD dilution trap before (e.g. de Baar et al. 2005), I refer to Smetacek and Naqvi (2008) and the comments to Smetacek et al. (2012) where the issue has been explained in detail. Reference to the effect of MLD on phytoplankton biomass is given only in Lines 9-12 of page 17 but the MLD values used in the model need to be explained explicitly in the text and, where appropriate, in the legends.

Only concentrations are referred to in this paper and the model results are compared with satellite images that measure concentration in the upper few metres. How the effects of much deeper MLDs in the SO as compared to the N. Pacific and N. Atlantic have been taken into account needs to be highlighted in the text and legend of figure 4.

MLD is explicitly produced by the dynamic ocean physical model (NEMO) based on the local atmospheric conditions as represented in forcing data from the ECMWF re-analysis. We clarified this in the text. More importantly, MLD is identical in both the PlankTOM10 and PlankTOM6 simulations, and thus the differences between these two models can only be attributed to the ecosystem structure. We focus here on surface concentrations because we are trying to explain the surface differences between the North and South as observed via satellite Chl data. We clarified in Fig. 4 and throughout the paper that we report the surface concentration for Chl and biomass for the top model box (10 m). These surface conditions reflect the internal dynamics of mixing created by changes in MLD. We further enhanced the results and discussion to mention the role of MLD in the Northern and Southern hemisphere.

The new text reads as:

Methods Section 2.4 (1st paragraph): "The model resolves 30 vertical levels, with a 10 m depth resolution in the upper 100 m. NEMOv3.1 calculates vertical diffusion explicitly and represents eddy mixing using the parameterisation of Gent and McWilliams (1990). The model thus generates its own mixed-layer dynamics and associated mixing based on local buoyancy fluxes and wind conditions."

Results Section 3.3 (2nd paragraph): "The failure of PlankTOM6 to reproduce the observed low Chl concentration in the Southern Ocean summer is further highlighted in Fig. 9, which shows the seasonal cycle of mean Chl for the Northern Hemisphere and the Southern Ocean, where it is most pronounced. In PlankTOM6, the seasonal cycle in the North and South are very similar, with the slightly lower concentrations in Southern Ocean summer possibly caused by a slightly deeper summer time mixed-layer depth (29m compared to 19m). In contrast in PlankTOM10, the seasonal cycle of Chl in the South is flatter and concentrations are always below those of the North, as in the observations. As PlankTOM6 and PlankTOM10 have identical physical environments (including mixed-layer depth), the North-South differences are entirely due to ecosystem structure. "

Discussion Section 4 (4th paragraph): "Our results indicate that zooplankton grazing plays an important control on Southern Ocean Chl. This importance propagates through to phytoplankton biomass. Indeed, the North/South surface ratio of phytoplankton biomass is also larger in PlankTOM10 (1.62) compared to PlankTOM6 (1.18), very close to the corresponding modelled North/South ratio of Chl. The difference between the two models also persists through depth until about 300 m where the biomass of both model approaches zero. Because of these marked differences it is clear that the representation of global biogeochemical cycles in ocean models will be influenced by the ecosystem composition. In both PlankTOM6 and PlankTOM10, the mesozooplankton and macrozooplankton faecal pellets aggregate into the same large, fast-sinking particles, limiting the influence of zooplankton of different sizes. To discern the effect of marine ecosystem composition on global biogeochemical cycles, a wider spectrum of particle size classes sinking at different speeds would be needed (e.g. Kriest; 2002). Such an improved vertical dynamics of the twilight zone, together with the enhanced representation of zooplankton dynamics presented in this paper, would allow further exploration of the interactions between iron fertilisation, grazing and mixed-layer dynamics, which have led to large differences among ocean iron fertilization experiments (Smetacek and Naqvi 2008; Boyd et al. 2008). "

In the model, meso- and macrozooplankton biomass is derived from 200 m vertical net tows and then converted to concentration in the 200 m water column (lines 12 ,Äì 15, page 12). Phytoplankton concentrations should be treated in an equivalent fashion but for the MLD and not a standard depth as for zooplankton. It is likely that the differences in surface chlorophyll concentrations between satellite and model will reduce further when this effect is considered.

Please note that in the model results, meso- and macrozooplankton biomass are presented for the surface ocean only, as is phytoplankton concentration, and thus all biomass metrics are consistent. Zooplankton biomasses are thus not derived from vertical net tows, but rather produced in situ by the model based on the growth and loss term equations presented in the Supplementary Material. Only the observed meso- and macrozooplankton biomasses presented in Figure 3 (panels d and f) are based on tows. We had already noted in the text that when comparing the observations with the model, the depth difference in the sampling method can explain a factor of 1.5 to 2.0 in biomass. Given the uncertainty in the data and the large regional spread in observed zooplankton biomass, it would bring little additional information to scale the data for the purpose of Figure 3 and could introduce additional errors. Instead, we have now explained the difference in the figure caption directly. Everywhere else throughout the manuscript only surface biomass is discussed. We now explain in the discussion (Section 4, see response above) why we do not analyse more the fluxes to depth and interactions with MLD any

further, which we intend to present in a subsequent analysis that will require further model development.

New text to Section 2.1 (3rd paragraph): "The PFT biomasses are produced by the model for each grid box based on the growth and loss term equations presented in Supplementary Material."

New text caption of Figure 3 reads: "All data are for the surface, generally corresponding to the mixed layer, except for Chl, which is seen by satellite over one optical depth, and mesozooplankton and macrozooplankton, which are from depth-integrated tows and may underestimate surface concentrations (by a factor 1.5-2; see text)."

b) Chlorophyll-biomass ratios Chlorophyll synthesis is one of many biochemical pathways that are limited by iron deficiency. Providing iron to phytoplankton leads to increased production of this pigment and C/chl ratios can drop two to threefold with only a marginal increase in biomass, an effect which is particularly apparent in diatoms. So chlorophyll concentrations are an unreliable proxy for phytoplankton biomass because they can vary so much between iron-limited and iron-replete conditions. The ratio (Chl/C) has been considered as a determinant of growth rate in the model but I could not ascertain whether the ,"phytoplankton biomass" based on chlorophyll in the SO was also corrected for this variable. This would exacerbate the chlorophyll problem by increasing the real, carbon-based biomass.

Indeed the ratio Chl/C varies as a function of Fe, as does the phytoplankton biomass. This is fully considered in the model, and was clarified in the text:

Methods Section 2.1 (5th paragraph): "Phytoplankton PFT growth rates are also limited by light and inorganic nutrients (P, N, Si, and Fe) using a dynamic photosynthesis model that represents the two-way interaction between photosynthetic performance and Fe/C and Chl/C ratios (Buitenhuis et al. 2013a)."

We now make an explicit link between Chl and phytoplankton biomass in the Discussion, which confirms that the North/South differences in Chl are also present in phytoplankton biomass (see new text in Section 4, 4th paragraph, mentioned above).

Question 2 Why only 3 zooplankton PFTs? What would the model results look like if a fourth zooplankton PFT that included the salps was introduced: micro-feeding, fast-growing, large zooplankton? Put in another way, why stop at the third category of zooplankton? They have many more PFTs than phytoplankton. Salps are relevant because their stocks have been increasing in the SO over the past decades concomitant with a krill decline (Atkinson et. al. 2004) so it is possible that a replacement is taking place with consequences for SO productivity. Salps differ from the category of zooplankton added here (slow-growing macrozooplankton) in that they have short generation times because they can produce individual animals by budding in the course of a day to weeks depending on temperature and presumably food supply. Their inclusion might lead to complete grazing down of the phytoplankton because nothing is known about the checks and balances on their population size. Since they swim, feed and breathe simultaneously it should be possible to model their grazing efficiency fairly easily using threshold values (particle concentration at which they starve, multiply, etc.) from the literature. If the data have not been collected yet, the model results could be used to generate interest in this question.

Indeed it would be very interesting to have additional zooplankton and we are working on new model versions that explicitly include pteropods, foraminifera and salps. However, these are substantial additions to the model that require a lot of time (one researcher and one PhD

student are working full time on this new model version). The model version presented here with 3 zooplankton is the simplest model that succeeds in reproducing the North/South ratio in Chl, which is already an advance compared to previous models and which we deem worthy of a publication on its own, i.e. each addition of a PFT to the model must be fully understood and evaluated by the scientific community before proceeding to further additions. We added in the conclusion that the addition of further zooplankton compartments could result in even more realistic horizontal distribution of Chl. The additional text in the discussion Section 4 (last paragraph) reads: "In addition, the model does not include several other ecosystem pathways known to play an important role, such as viral lysis (Evans et al., 2009), and the zooplankton representation does not include salps, pteropods, and auto- and mixotrophic dinoflagellates. The nano- and microzooplankton are also combined into a single compartment."

Question 3: Has iron recycling due to zooplankton grazing been considered in the model? The authors refer to "the dynamics of the SO zooplankton community" as being a more important determinant of low summer phytoplankton biomass than iron limitation. They mention that trophic cascades within the zooplankton have been built into the model but there is no mention of recycling of the limiting nutrient, in this case iron, by the zooplankton, i.e. a feedback loop which would stimulate net production. At the end of the abstract and in the conclusions one is reminded that zooplankton make fast-sinking faecal pellets and carry out vertical migration, implying that the more zooplankton, the more vertical flux and carbon sequestration. So, in today's times it might be a good thing to have zooplankton around because they sink carbon, a few decades ago, before collapse of the traditional commercial fisheries, zooplankton were good because they made fish food. Since the zooplankton category introduced here are long-lived, their food supply would have to be sustainable, so one wonders what percentage of the faecal pellets produced sinks out of the mixed layer. If all were to sink out, the surface layer would soon be depleted of essential elements and the grazers would starve; so the category introduced here only evolved because most of its wastes are recycled in the surface layer. Salps on the other hand are roving grazers that can afford to let their wastes sink behind them. This is just to mention a few complications that arise when attempting to model the impact of zooplankton and nekton on the biogeochemistry of the surface layer.

Yes, iron recycling due to zooplankton grazing is considered in the model. We now specify this in the Methods Section 2.1 (4th paragraph): "The model keeps track of the Chl, Si and Fe content of organic material, which are released when organic material is remineralised following bacterial respiration and zooplankton grazing." Our iron model is parameterised according to Aumont and Bopp (2006), and it successfully reproduces the changes in surface Chl concentration following iron enrichment as noted in Section 3.4.3)."

General comments In order to continue improvement of biogeochemical models by exploring the impact of zooplankton dynamics on ecosystem structure and functioning, it will be necessary to develop a framework of interactions based on the evolutionary ecology of the phytoplankton/zooplankton relationship. Unfortunately this has not received the attention it deserves for various reasons (see Smetacek et al. 2004, Smetacek 2012), in particular, because the necessary, dependable, quantitative information is still lacking. To my mind this can only come from studies carried out in situ because enclosures of any sort will hamper the zooplankton and nekton. Furthermore, comprehensive measurement programmes of the same body of water will be necessary to study rates and processes within functioning ecosystems. Perturbing the system under study in order to identify shifts in the mechanisms would enhance the value of the information gained from sustained measurements of the same water mass. The

whole-lake experiments carried out in the 1970s in the USA and Canada brought unsuspected trophic cascades into focus: Lakes changed their colour depending on the presence or absence of predatory fish. "Were it not for whole lake experiments, limnology would be where bio-oceanography is today, firmly entrenched in the bottom-up paradigm. It follows that bio-oceanography could be where limnology is today if more dedicated in situ experiments are carried out by the scientific community" (comments in Smetacek et al 2012). Could removal of the whales, that once lived sustainably from krill biomass equivalent to double the global commercial fish catch, have had an effect on SO chlorophyll concentrations? Such hypotheses could be tested with ocean iron fertilization (OIF) experiments that have proven themselves as the equivalent of whole-lake experiments in the sea.

We would welcome this kind of experimental data to provide mechanistic information on the phytoplankton/zooplankton interactions. We added a reference to Smetacek et al. 2004 in the conclusions Section 5 (1st paragraph).

All ocean iron fertilization (OIF) experiments carried out so far have shown that iron addition led to a substantial increase in the photosynthetic efficiency index (Fv/Fm) and remained higher for many weeks inside the fertilized patch than values measured concomitantly in surrounding iron-limited water. These results, bolstered with direct measurements of iron concentrations and various types of bottle experiments have unequivocally shown that phytoplankton growth rates in HNLC areas are limited by the iron supply. The conclusion of this paper: that zooplankton grazing rather than iron controls phytoplankton biomass build up would imply, framed in John Martin's iron hypothesis, that fluctuations in the grazer populations were responsible for climate cycles. This is probably not what the authors mean so the wording of their concluding remarks needs to be properly qualified rather than presented in a simplistic sweeping statement.

Please refer to Section 4 last paragraph which covered much of the qualified language that is called for in this comment. We have now moved this paragraph to the end of the Conclusions, where it has more weight and visibility. It reads: "Our results on the important role of grazing do not contradict the results on the importance of Fe-fertilisation as highlighted in Fe enrichment experiments (Boyd and al., 2007), because additional Fe will trigger further growth provided that Fe is initially below an optimal concentration (Blain et al., 2007). However, our results suggest that low Fe concentrations by themselves are insufficient to account for the very low Chl levels observed in the Southern Ocean HNLC region in summer, and that differences in zooplankton trophic and community structure, and concomitant grazing dynamics have an important role in controlling phytoplankton blooms and maintaining very low Chl levels in that region. Although previous studies emphasised the role of phytoplankton community structure (Arrigo et al., 1999) and mixed layer dynamics for nutrient supply and demand (Platt et al., 2003a;Platt et al., 2003b) in ocean biogeochemical cycles, our analysis makes it clear that it is important to consider the whole pelagic ecosystem, including the zooplankton, when studying and predicting ecosystem responses to Fe (or any essential nutrient) fertilisation. This complex interplay has received less attention than either the drivers of primary production or the representation of Fe cycling in global biogeochemical modelling. Our results suggest that representing zooplankton interactions more explicitly could lead to more mechanistic representation of biogeochemistry – climate interactions. "

While we do not want to imply that zooplankton grazing is responsible for glacial-interglacial CO_2 variations, we note here that the current consensus is that iron fertilisation caused no more than 1/3 of the observed glacial-interglacial atmospheric CO_2 variations and that global

biogeochemistry models still fail to reproduce the other 2/3 of the observed atmospheric signal. Therefore there are clearly important issues to resolve in the next generation of global biogeochemical models. We cannot show here whether zooplankton had an influence on atmospheric CO_2 levels on millennial time scales, but we cannot rule it out given their importance in determining the vertical flux and regional distribution of carbon in the ocean.

Nevertheless, the improved model presented here clearly demonstrates that increasing zooplankton PFTs is an important way to nudge biogeochemical models closer to reality. The last sentences of the Conclusions focus on the possible outcome of ocean iron fertilization (OIF): "Assessments of the impact of such geoengineering will be unreliable, at least until the full ecosystem response including the grazing pathways (Landry et al., 1997) and the relationship to deep water carbon export (Smetacek et al., 2012) can be reproduced with models, which could be used to make quantitative predictions". The authors are putting the cart before the horse here: OIF experiments are the most reliable way of achieving the assessments called for here. So far only puddles of a few weeks' duration have been studied. Larger, longer-term experiments are needed to assess the impact of higher trophic levels on ocean biogeochemistry. The ensuing model would then permit one to extrapolate from the experimental scale to the really large-scale of relevance to the global carbon cycle, should this be deemed necessary.

Indeed, OIF may be the most reliable, but they are local in time and space by nature. The only way to make a large-scale assessment of the quantitative potential of purposeful Fe-fertilisation is to use models to scale up the observations. We have specified in the conclusion Section 5 (3rd paragraph) that our comment on the role of models refer to making quantitative predictions "over large areas".

References

Smetacek, V. Making sense of ocean biota: How evolution and biodiversity of land organisms differ from that of the plankton. J. Biosci. 37, 589-607, DOI 10.1007/s12038- 012-9240-4, 2012 Smetacek, V., Assmy, P., Henjes, J.: The role of grazing in structur- ing Southern Ocean pelagic ecosystems and biogeochemical cycles, Antarctic Science 16: 541-558, 2004 Smetacek, V. and Naqvi, S.W.A.: The next generation of iron fertilization experiments in the Southern Ocean. Phil Trans. R. Soc. A 366, 3947-3967, 2008.